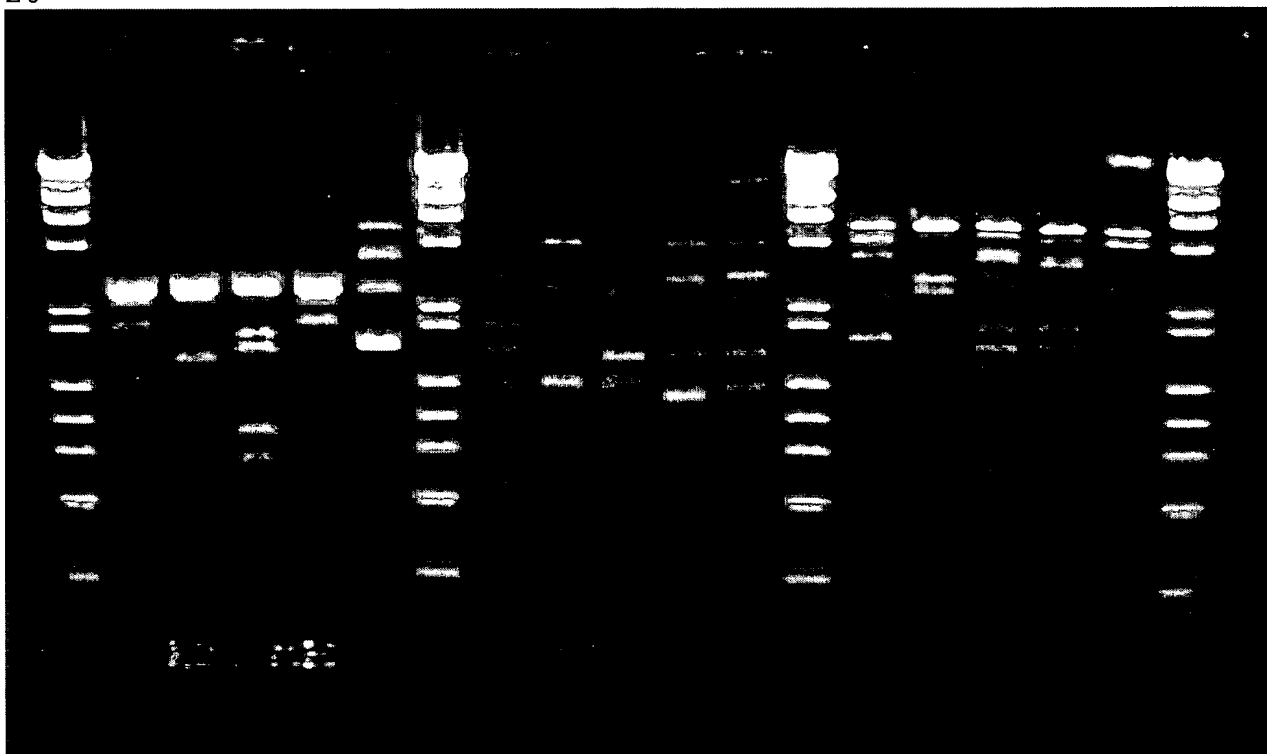


CstMI Figure 1 - Agarose gel showing CstMI cleavage of
lambda, T7, phiX174, pBR322 and pUC19 DNAs.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
20



CstMI Figure 2 - DNA sequence of the CstMI gene locus (SEQ ID NO:).

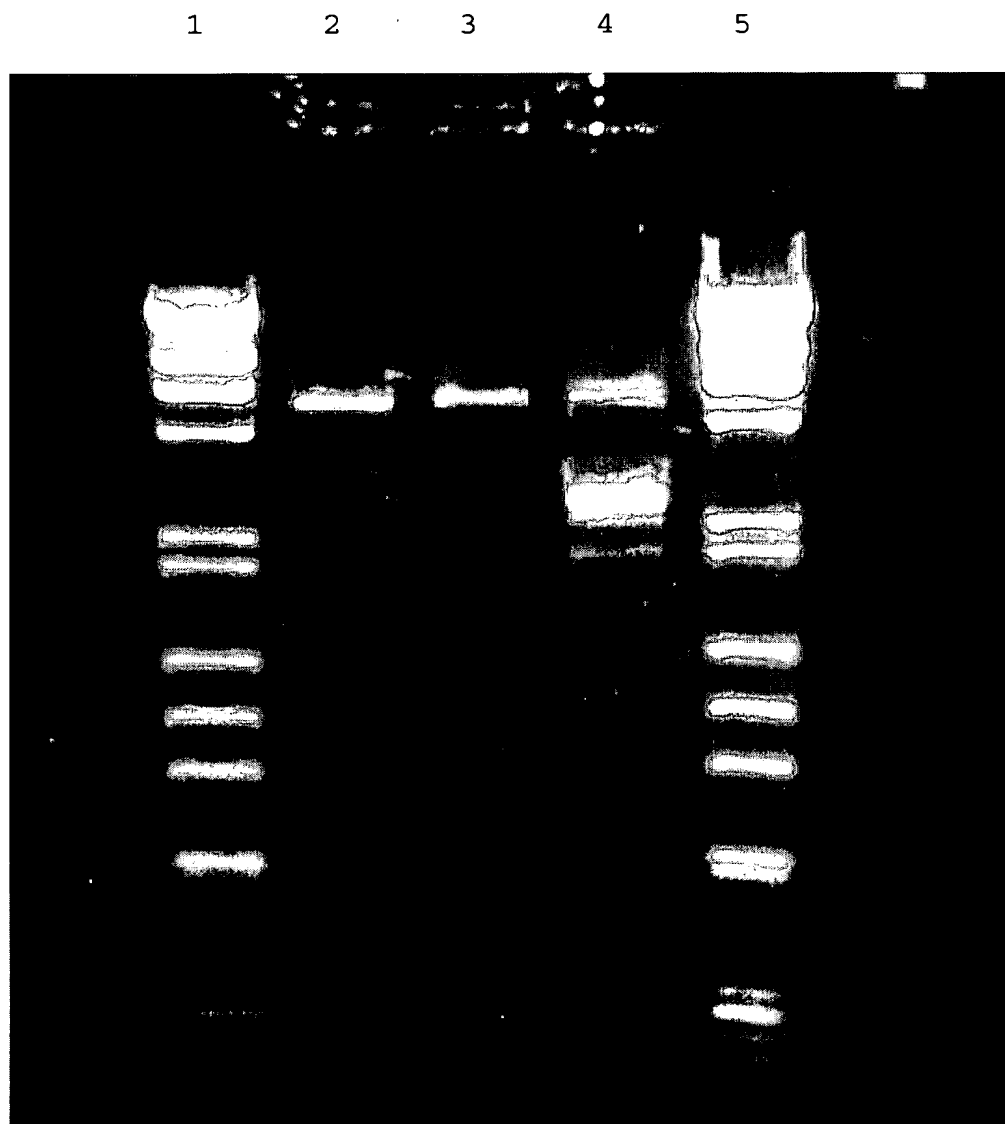
```
1  ATGGTTATGG CCCCTACGAC TGTTTTTTGAC CGCGCTACCA TTCGCCACAA
51  TCTCACCGAA TTCAAACCTCC GGTGGCTTGA CCGCATTAAG CAATGGGAGG
101 CGGAAAACCG ACCCGCAACC GAGTCGAGTC ACGACCAACA GTTCTGGGGT
151 GACCTGCTCG ACTGCTTCGG TGTCAACGCC CGCGACCTGT ACTTGTACCA
201 ACGCAGCGCT AAACGCGCTT CGACGGGGCG CACCGGCAAG ATCGACATGT
251 TTATGCCGGG CAAAGTCATA GGCGAGGCTA AGTCCCTCGG CGTCCCGCTC
301 GATGATGCTT ATGCCCAAGC TTTGGATTAT TTGCTGGGCG GTACTATCGC
351 GAACTCGCAC ATGCCGGCCT ATGTTGTCTG CTCCAACCTC GAGACCTTGC
401 GGGTTACCCG TCTTAACCGC ACCTATGTCG GCGATAGCGC CGACTGGGAC
451 ATTACATTCC CTTTAGCTGA GATTGACGAG CACATCGAAC AACTCGCTTT
501 TCTCGCCGAC TATGAAACCT CCGCCTACCG GGAGGAAGAA AAGGCTTCCC
551 TGGAAAGCCTC TCGGTTAATG GTGGAGCTCT TCCGCGCCAT GAACGGCGAC
601 GACGTGGACG AGGCAGTAGG CGATGACGCT CCCACCACGC CGGAGGAAGA
651 AGACGAGCGC GTCATGCGCA CCTCTATCTA CCTCACCCGA ATCCTCTTCC
701 TTCTCTTCGG CGACGACGCA GGACTCTGGG ATACCCCGCA TTTGTTTGCG
751 GACTTTGTGC GCAATGAAAC CACCCAGAA TCGCTCGGCC CGCAGCTCAA
801 TGAGCTATTT AGCGTGCTTA ATACCGCCCC GGAAAAGCGG CCTAAGCGTT
851 TGCCATCAAC GTTGGCGAAG TTTCTTATG TCAATGGTGC CCTATTTGCT
901 GAACCGTTGG CCTCGGAGTA CTTCGACTAC CAGATGCGCG AAGCATTTGCT
951 TGCTGCCTGC GACTTCGACT GGTGACCAT TGACGTCTCC GTCTTTGGTT
1001 CGTTGTTCCA ATTGGTGAAA TCGAAGGAAG CGCGCCGCAG CGACGGCGAA
1051 CACTACACGT CTAAGGCCAA CATCATGAAG ACCATCGGCC CGCTGTTTTT
1101 GGACGAGCTG AGGGCTGAGG CCGATAAGTT GGTGTCTTCT CCGTCGACGT
1151 CGGTGGCCGC ATTAGAGCGC TTCCGCGACT CCCTGTCTGA GCTGGTATTC
1201 GCTGATATGG CTTGTGGTTC TGGAAACTTC CTGCTTCTGG CGTATCGGGA
1251 GTTGCGCCGG ATTGAAACCG ACATCATTGT CGCTATACGC CAGCGCCGCG
1301 GTGAAACGGG CATGTCGTTG AATATTGAGT GGGAGCAGAA ACTGTCCATT
1351 GGGCAGTTCT ACGGCATTGA GCTGAATTGG TGGCCTGCCA AGATTGCTGA
1401 GACTGCCATG TTCCTAGTTG ACCATCAGGC CAACAAGGAG CTTGCCAACG
1451 CTGTGGGTTAG GCCTCCGGAG CGGTTGCCGA TTAAGATTAC CGCGCACATT
1501 GTGCACGGCA ATGCCCTGCA GCTTGATTGG GCAGACATAC TCTCGGCTTC
```

1551 TGCCGCCAAG ACGTATATCT TCGGTAACCC GCCGTTTTTG GGGCATGCGA
1601 CGAGAACTGC TGAACAAGCT CAAGAACTCC GAGACTTGTG GGGCACTAAG
1651 GACATTTTAC GCTTGGACTA CGTCACCGGC TGGCATGCAA AGTGCTTGGA
1701 TTTCTTTAAG TCCCGAGAGG GTCGTTTTGC GTTTGTCACC ACCAATTCAA
1751 TTACTCAAGG TGATCAAGTT CCACGGCTAT TTGGGCCTAT CTTCAAAGCA
1801 GGGTGGCGTA TTCGTTTCGC TCACCGCACG TTTGCGTGGG ACTCTGAAGC
1851 ACCCGGTAAA GCTGCTGTTC ACTGCGTCAT TGTGGCTTC GATAAGGAGA
1901 GTCAACCACG TCCACGTCTG TGGGATTATC CCGATGTAAA GGGCGAGCCA
1951 GTCTCAGTGG AAGTAGGCCA GTCCATTAAT GCCTATTTAG TAGACGGCCC
2001 TAATGTTCTT GTCGATAAAT CCCGGCATCC TATTTGTCG GAAATATCGC
2051 CCGCAACTTT TGGAAATATG GCGCGAGATG GCGGCAACCT TCTAGTTGAG
2101 GTCGACGAAT ACGACGAGGT TATGAGTGAC CCCGTAGCGG CAAAGTATGT
2151 TCGCCCTTTC CGGGGTAGTC GAGAGCTAAT GAACGGCTTA GATCGGTGGT
2201 GTCTATGGCT TGTAGATGTA GCACCGTCAG ACATTGCCCA GAGTCCGGTT
2251 CTGAAAAAGC GTCTAGAAGC GGTAAAGTCT TTTGAGCCG ACAGTAAAGC
2301 GGCAAGTACA CGGAAAATGG CTGAAACTCC GCACTTATTC GGCCAGCGGT
2351 CGCAACCGGA TACTGATTAC CTTTGCCTGC CGAAGGTAGT AAGCGAACGC
2401 CGCTCGTATT TCACCGTACA AAGGTATCCA TCAAACGTAA TCGCTTCTGA
2451 CCTAGTATTC CATGCTCAAG ATCCAGACGG CCTGATGTTT GCGCTAGCGT
2501 CGTCGTCGAT GTTCATTACG TGGCAGAAAA GCATCGGAGG ACGACTCAAG
2551 TCTGATCTCC GTTTTGCTAA CACTTTGACG TGAATACTT TCCCAGTGCC
2601 AGAACTCGAC GAGAAGACGC GGCAGCGAAT TATTAAAGCG GGCAAGAAGG
2651 TGCTCGACGC CCGCGCGCTG CACCCAGAAC GCTCGCTGGC CGAGCACTAC
2701 AACCCTCTCG CGATGGCACC GGAAGTCATC AAAGCGCATG ATGCGCTCGA
2751 CCGCGAGGTG GATAAAGCGT TTGGCGCGCC ACGAAAGCTG ACAACTGTTC
2801 GGCAGCGCCA GGAGCTATTG TTTGCCAATT ACGAAAACT CATCTCACAC
2851 CAGCCCTAG

CstMI Figure 3: Amino acid sequence of the CstMI gene locus (SEQ ID NO:).

```
1  MVMAPTTVFD RATIRHNLTE FKLRLDRIK QWEAENRPAT ESSHDQQFWG
51  DLLDCFGVNA RDLYLYQRSA KRASTGRTGK IDMFMPGKVI GEAKSLGVPL
101 DDAYAQALDY LLGGTIANSN MPAYVVCNSF ETLRVTRLNR TYVGDSADWD
151 ITFPLAEIDE HIEQLAFLAD YETSAYREEE KASLEASRLM VELFRAMNGD
201 DVDEAVGDDA PTTPEEEDER VMRTSIYLTR ILFLLFGDDA GLWDTPHLFA
251 DFVRNETTPE SLGPQLNELF SVLNTAPEKR PKRLPSTLAK FPYVNGALFA
301 EPLASEYFDY QMREALLAAC DFDWSTIDVS VFGSLFQLVK SKEARRSDGE
351 HYTSKANIMK TIGPLFLDEL RAEADKLVSS PSTSVAALER FRDSLSELVF
401 ADMACGSGNF LLLAYRELRR IETDIIVAIR QRRGETGMSL NIEWEQKLSI
451 GQFYGIELNW WPAKIAETAM FLVDHQANKE LANAVGRPPE RLPIKITAHI
501 VHGNALQLDW ADILSASAAK TYIFGNPPFL GHATRTAEQA QELRDLWGTK
551 DISRLDYVTG WHAKCLDFFK SREGRFAFVT TNSITQGDQV PRLFGPIFKA
601 GWRIRFAHRT FAWDSEAPGK AAVHCVIVGF DKESQPRPRL WDYPDVKGEP
651 VSVEVGQSIN AYLVDGPNVL VDKSRHPISS EISPATFGNM ARDGGNLLVE
701 VDEYDEVMSD PVAACYVRPF RGSRELMNGL DRWCLWLVDV APSDIAQSPV
751 LKKRLEAVKS FRADSKAASR RKMAETPHLF GQRSQPDYDY LCLPKVVSER
801 RSYFTVQRYP SNVIASDLVF HAQDPDGLMF ALASSSMFIT WQKSIGGRLK
851 SDLRFANTLT WNTFPVPELD EKTRQRIIKA GKKVLDARAL HPERSLAEHY
901 NPLAMAPELI KAHDALDREV DKAFGAPRKL TTVRQRQELL FANYEKLISH
951 QP
```

Figure 4 - Agarose gel showing CstMI protection of pTBCstMI.3 DNA and cleavage of unmodified DNA substrate.



CstMI Figure 5 : Determination of the CstMI cleavage site.

Figure 5A: Location of cleavage on 5'-AAGGAG-3' strand.

pUC19-Adeno2BC4 DNA was cut with CstMI producing ends as indicated by the arrows:

5'--..CGAACCCAGGTGTGCGACG↓TCAGACAACGGGGGAGCG**CTCCTT**TTG..-
3'

(SEQ ID NO:3)

3'--..GCTTGGGTCCACACGCT↑GCAGTCTGTTGCCCCCTCGC**GAGGAA**AAC..-
5'

The resulting cleaved DNA:

5'--..CGAACCCAGGTGTGCGACG-3' (SEQ ID NO:4)

3'--..GCTTGGGTCCACACGCT-5'

The template strand for dideoxy DNA sequencing extension:

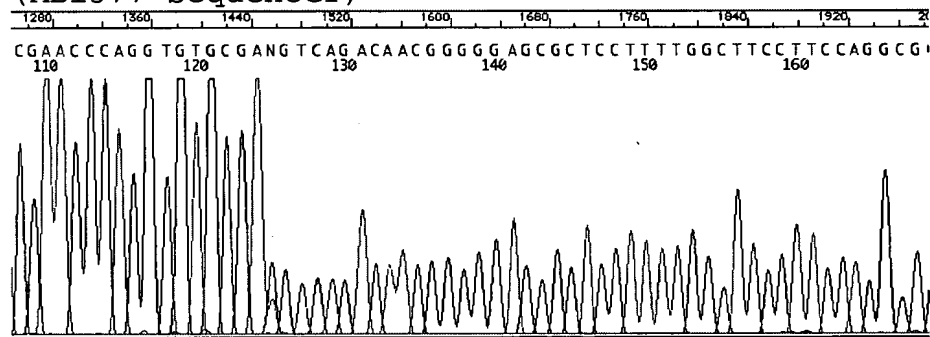
3'--..GGGTCCACACGCT-5'

The primer (NEB1224) is annealed and extended through the CstMI site. When the reaction reaches the end of the molecule the Taq polymerase adds an extra A base.

5'-PRIMER->. . CGAACCCAGGTGTGCGA(A)-3' (SEQ ID NO:5)

3'-. GCTTGGGTCCACACGCT-(N20-GAGGAA)-5'

Sequencing Profile of CstMI cut pUC19-Adeno2BC4 DNA
(ABI377 Sequencer)



CstMI Figure 5: Determination of the CstMI cleavage site.

Figure 5B: Location of cleavage on 5'-CTCCTT-3' strand.

pBR322 DNA was cut with CstMI, yielding ends indicated by the arrows:

5'--..TGCATGC**AAGGAG**ATGGCGCCCAACAGTCCCCC↓GGCCACGGGGCC..-
3'

(SEQ ID NO:6)

3'--..ACGTACG**TTCCTC**TACCGCGGGTTGTCAGGG↑GGCCGGTGCCCCGG..-
5'

The resulting cleaved DNA:

5'--..TGCATGC**AAGGAG**ATGGCGCCCAACAGTCCCCC -3'

(SEQ ID NO:7)

3'--..ACGTACG**TTCCTC**TACCGCGGGTTGTCAGGG -5'

The template strand for dideoxy DNA sequencing extension:

3'--..ACGTACG**TTCCTC**TACCGCGGGTTGTCAGGG -5'

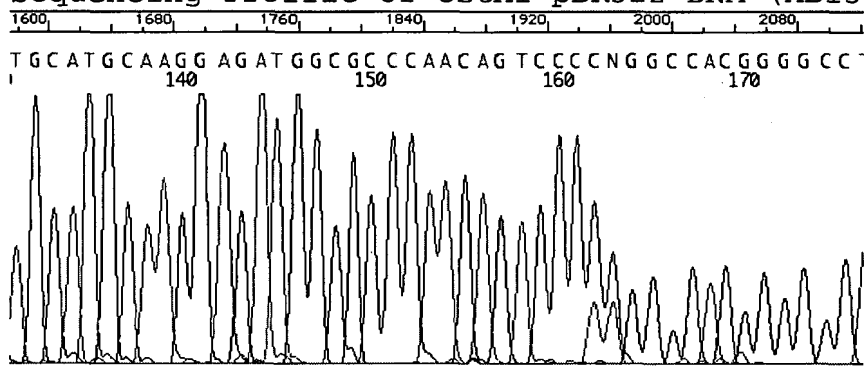
The primer (NEB1242) is annealed and extended through the CstMI site. When the reaction reaches the end of the molecule the Taq polymerase adds an extra A base.

5'-PRIMER->. TGCATGC**AAGGAG**ATGGCGCCCAACAGTCCC(A)-3'

(SEQ ID NO:8)

3'--. ACGTACG**TTCCTC**TACCGCGGGTTGTCAGGG -5'

Sequencing Profile of CstMI pBR322 DNA (ABI377 Sequencer)



CstMI Figure 6: Sequence alignment of CstMI and MmeI amino acid sequences.

Symbol comparison table: /gcg/bin/gcgcore/data/rundata/blosum62.cmp
CompCheck: 1102

Gap Weight:	8	Average Match:	2.778
Length Weight:	2	Average Mismatch:	-2.248
Quality:	1548	Length:	942
Ratio:	1.718	Gaps:	19
Percent Similarity:	51.009	Percent Identity:	39.574

Match display thresholds for the alignment(s):

| = IDENTITY
: = 2
. = 1

CstMI.pep x MmeI.pep June 20, 2003 11:45 ..

```
20 EFKLRLWLDRIKQWEAENRPATESSHDQQFWGDLDCFGVNARDLYLYQRS 69
| : : : : | . | | . | . | | : | | : : . : : .
7 EIRRKAIEFSKRWE...DASDENSQAKPFLIDFFEVFGITNKR VATFEHA 53

70 AKRASTGR....TGKIDMFMPGKVI GEAKSLGVPLDDAYA QALDYLLGGT 115
| : . | : : | | | . : | | | | | | | | | | | |
54 VKKFAKAHKEQSRGFVDLFWPGILLIEMKSRGKDL DKAYDQALDYFSG.. 101

116 IANSHMPAYVVC SNFETLRVTRLNRTYVGDSADWDITFPLAEIDEHIEQL 165
| | : | | | . . | : | . | | : | : | | : : :
102 IAERDLPRYVLVCD FQRFLTDL...ITKES....VEFLLKDL YQNVR SF 144

166 AFLADYETSAYREEEKASLEASRLMVELFRAMNGDDVD EAVGDDAPT TPE 215
| : | | : | : : . : . | . | | | :
145 GFIAGYQTQVIKPQDPINIKAAERMGKL.....HDTLKL VGYEGHA... 185

216 EEDERVMRTSIY LTRILFLLFGDDAGLWDTPHLFADFVRNETTPE..SLG 263
: | | | : | | | | : | : : | | : : : . | : |
186 .....LELYLVRL LFCLFAEDTTIFE.KSLFQ EYIETKTLEDGSDLA 226

264 PQLNELFSVLNTA PEKRPKRLPSTLAKFPYVNGALFAEPLASEYFDYQMR 313
: | | | | | : | | | | | | | | | | | | | | | |
227 HHINTLFYVLNTPEQKRLKNLDEHLA AFPYINGKLFEEPLPPAQFDKAMR 276

314 EALLAACDFDWST IDVSVFGSLFQLVKSKEARRSDGEHYTSKANIMKTIG 363
| | | | | | | | | | : | | | | | : . | | . | | | | | : | |
277 EALLDLCSLDWSR ISPAIFGSLFQSIMDAKKRRNLGAHYTSEANILKLIK 326

364 PLFLDELRAEADKL VSSPSTSVAALERFRDSLSELVFADMACGSGN FLLL 413
| | | | | | | | : | . . . . . | | | | | | | | | | :
327 PLFLDELWVEFEK VKNKNKNKLLA....FHKKLRGLTFFDPACGCGN FLVI 372

414 AYRELRR IETDIIVAIRQRRGETGMSLNIEWEQKLSIGQFYGIELNWWPA 463
| | | | : | : : : | | : | . | | . : : : | | : | | : . | |
373 TYRELRLLEIEV LRGL.HRGGQ..QVLDIEHLIQINVDQFFGIEIEEFPA 419

464 KIAETAMFLVDHQA NKELANAVGRPPERLPIKITAHIVHGNALQLDWADI 513
. | | : | : . | | | | . . . . . | : : | | | | . | | | : | |
420 QIAQVALWLT DHQNMNMKISDEFGNYFARIPLKSTPHILNANALQIDWNDV 469

514 LSASAAKTYIFGN PPF LGHATRTAEQAQELRDLWGT.KDISRLDYVTGWH 562
| | : | | | | . | . | | : | . . | | | | | | | :
```


470 LEAKKC.CFILGNPPFVGKSKQTPGQKADLLSVFGNLKSASDLDLVAAWY 518
 563 AKCLDFFKSREG.RFAFVTNSITQGDQVPRLFGPIFKAGWRIRFAHRTF 611
 | : .. | |||.|||||:| |. : | :| |||||
 519 PKAAHYIQTNANIRCAFVSTNSITQGEQVSLWPLLLSLGIKINFAHRTF 568
 612 AWDSEAPGKAAVHCVIVGFDKESQPRPRLWDYPDVKGEPVSVEVGQSINA 661
 .| .|| | |||||:| | . :|| : |||... ..||
 569 SWTNEASGVAAVHCVIIGFGLKDSDEKIIYYESINGEPLAIK.AKNINP 617
 662 YLVDGPNVLVDKSRHPISSEISPATFGNMARDGGNLLVEVDEYDE.VMSD 710
 || || .|: | . || |.: :|| | || | :| .: : .:
 618 YLRDGVVDVIACKRQQPI.SKLPSMRYGNKPTDDGNFLFTDEEKNQFITNE 666
 711 PVAACYVRPFRGSRELMNGLDRWCLWLVDVAPSDIAQSPVLKKRLEAVKS 760
 | . || | | | | .| ||||| :| |.. |:| .
 667 PSSEKYFRRFVGGDEFINNTSRWCLWLDGADISEIRAMPLVLARIKKVQE 716
 761 FRADSKAASTRKMAETPHLFGQRSQPDYDYLCLPKVVSERRSYFTVQRYP 810
 || | | ||. | || | |||||:|. || | : :
 717 FRLKSSAKPTRQSASTPMKFFYISQPDYDYLIPETSSENRQFIPIGFVD 766
 811 SNVIASDLVFHAQDPDGLMFALASSSMFITWQKSIGGRLKSDLRFANTLT 860
 |||.|. :| :|. | |||. | :. |||. | |: .|
 767 RNVISSNATYHIPSAEPLIFGLLSSTMHNCWMRNVGGRLESRYRYSASLV 816
 861 WNTFFVPELDEKTRQRIIKAGKKVLDARALHPERSLAEHYNPLAMAPELI 910
 :||| :. || . | . :| ||. :| ||| |. | ||:
 817 YNTFFWIQPNEKQSKAIEEAAFAILKARSNYPNESLAGLYDPKTMPSSELL 866
 911 KAHDALDREVDKAFGAPRKLTTVRQRQELLFANYEKLISHQP 952
 ||| ||: || :| :| || |:| :| |
 867 KAHQKLDKAVDSVYGFKGPNTET.ARIAFLEFETYQKMTSLLP 907

FIGURE 7

